

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Procedia Computer Science 5 (2011) 480–488



The 2nd International Conference on Ambient Systems, Networks and Technologies  
ANT-2011

## On Maximizing the Entropy of Complex Networks

Maytham H. Safar<sup>a\*</sup>, Ibrahim Y. Sorkhoh<sup>a</sup>, Hisham M. Farahat<sup>a</sup>, Khaled A. Mahdi<sup>a</sup>

<sup>a</sup>Kuwait University, P.O. Box 5969, Safat Kuwait

---

### Abstract

This work explores the configurations of complex networks that would exhibit the maximum entropy using both degree and cyclic entropies measures. Theoretical models for such networks were proposed and were validated by using our developed genetic algorithms. A complex network configuration with the maximum cyclic entropy was identified as a cyclic star network; a star network with all non central nodes is connected with a ring. Also GA results within a small statistical error conclude the existence of an irregular network of size 8 that has higher cyclic entropy than the cyclic star. A complex network with a uniform distribution topology has the maximum degree entropy as expected is identified using a deterministic algorithm.

© 2011 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

Selection and/or peer-review under responsibility of Prof. Elhadi Shakshuki and Prof. Muhammad Younas.

Keywords: Entropy; Topology; Complex Network

---

### 1. Introduction

In the search for the most robust and stable complex networks, we resolve to the fundamental problem of statistical mechanics; the problem of system equilibrium, continuous and discrete alike. Equilibrium is the state at which the system's entropy is the highest. In statistical mechanics, entropy is related to the number of system's internal states (microscopic states), given by the fundamental relation by Boltzmann,  $S = k \ln W$ , where  $S$  is the system's entropy,  $W$  is the total number of states and  $k$  is a constant. The total number of states is the number of configurations in the system as well. In the case of complex network, it is the number of possible topologies that a network can have.

The topology of a network can be defined in terms of the link and nodes in the network; such definition is simple, obvious and consistent with the direct relationship existing between nodes. Another definition is in term of triad, a cycle of degree 3, such definition is a coarse-grain of the former definition, and lesser

\* Maytham Safar. Tel.: +965-24987962

E-mail address: [maytham.safar@ku.edu.kw](mailto:maytham.safar@ku.edu.kw)

system's degree of freedom is needed for system representation. In addition, it introduced an ignored property in the former representation that is the feedback. In triads, feedback is taken care of however within a subset of three. The total number of possible configurations using triads' representation should lead to triadic entropy of the network. A more generalized definition of network topology is to consider all cycle sizes. The motivation is the inclusion of all possible feedbacks in the network. The fact that the sum of the all data bits in the network is conserved can accurately be represented in a set of cycles than in a set of links or triads. The total number of possible topologies using cycle's representation leads to the introduction of cyclic entropy.

Following a previous analysis, the robustness of a network, its ability to adapt to changes, is correlated to its ability to deal with internal feedbacks within the network. The more feedback loops (cycles) exist in the network, the more robust the network is. It follows that a fully connected network should represent the most robust network. Among sets of robust networks, the most stable robust network is a network of size 7 because it has the highest cyclic entropy among all sizes. Such interesting results motivate us to explore and search other network topologies that have the maximum cyclic entropy.

## 2. Related Works

Most of the complex systems in real life are represented by networks. Hence, studying network building methods for complex networks is of an increasing importance. Food webs, the Internet, protein-protein interactions, telephone calls, actor's network, social networks and many other systems are now represented by scientist from different fields as complex networks. The graph theory with its wide range of representations simplifies the study of such systems. Classification of the networks was one of the first major concerns in the complex networks field. One classification criteria suggests classifying a network depending on the (conceptually and computationally) simple properties of the network, nodes degrees, shortest paths and clustering coefficients. By applying this classification to the real life, the criteria lead to three different types of networks, scale-free, small-world, and random networks [1-3]. However, the methods used here were not enough to classify those networks accurately. In [4], the authors showed that this classification has enough ambiguity to be replaced. The new proposed classification of complex networks depends on the cycles' length distribution of the networks.

Another concern for scientists in studying complex network was how to evaluate the network robustness and its ability to tolerate changes. In [5, 6] they emerged the concept of network degree heterogeneity as a measurement of a networks' robustness. However, the result showed that telling whether or not a network is heterogeneous is not related to the networks' ability to face external attacks. Increasing the network heterogeneity will increase the vulnerability to intentional node attacks and at the same time it will decrease the vulnerability to random ones.

This concern and others urged the scientists to leverage a term that is widely used in thermodynamics, the Entropy. Entropy in thermodynamic is practically used to measure the systems' efficiency while conceptually it is related mainly to the system degree of chaos and the system ability to posses new states. Boltzmann Generic Entropy definition allowed scientists from different fields to use this concept to study systems robustness. As a first try, while he was expanding the information theory, Shannon was the first scientist who utilized the entropy principle to characterize communication systems. He used the entropy to calculate how much a communication systems' next states are unknown [7]. Another application of the entropy principle was in [2, 5]. By defining the systems' state as the degree distribution of the network, they used entropy to measure the systems' degree heterogeneity, hence the network robustness as they define it.

Another usage of entropy, the authors of [8, 9] suggested to use the entropy and the cycles distribution of social networks to measure the network robustness. The results showed that different samples of social

networks of different sizes have almost the same cycle's distribution. Another interesting result is that the maximum entropy of fully connected networks occurs with a network with size 7.

In this work, we are investigating the topologies of a general complex network that will cause the system to have the maximum degree entropy and cyclic entropy. The definition of such a system was defined in Shannon's' communication theory. The system that has all micro-states probabilities to be equally likely [10]. Thus the degree entropy will occur when all degrees have the same probability to occur and with the same manner, cyclic entropy will be maximum when the network has a uniform cyclic distribution.

### 3. Theory

Let  $\mathbf{k}$  be a set of discrete random variable that takes the following values  $\mathbf{k} = \{1, \dots, k, \dots, N\}$  with probabilities  $\mathbf{p} = \{P(1), \dots, P(i), \dots, P(N)\}$  respectively such that  $P(k) \geq 0$  and  $\sum_{k=1}^N P(k) = 1$ . There exists a measure of randomness, heterogeneity and uncertainties known *entropy*,  $H$  defined as

$$H(\mathbf{p}) = - \sum_{k=1}^N P(k) \ln(P(k)) \quad (1)$$

In a complex network context,  $P(k)$  may represent the degree distribution of links or the remaining degree (outward links) or cycles of size  $k$  in the network.

Degree distributions of links are the most common representation of  $P(k)$  in the literature. Different models of networks such as random, small world, scale free, exponential, and uniform and many others are usually represented and constructed as degree distribution models of actual networks; such as software, social, biological, circuits ...etc network. Simple degree distribution describes the connectedness of the network; hence the entropy will be a measure of heterogeneity, uncertainty of network connectedness.

The authors in [11] were the first to suggest computing the entropy based on the remaining (outward) degree distribution. The use of remaining degree distribution allows the calculation of degree-degree correlation and hence measures the network assortativeness. Therefore, the entropy should reflect the assortativeness uncertainty of the network.

The work in [8] suggests calculating the entropy based on the degree of cycles instead of links. Such approach emphasizes the feedback nature in complex network and indirectly measures the network ability to store information within the network cycles. Therefore, the cyclic entropy will measure the uncertainties associated with the information feedback in the network. One interesting finding about the degree of cycle's distribution is all networks have Gaussian (normal) distribution, each network differs in the value of variance and mean.

In all cases, the minimum entropy of the network  $H_{min} = 0$  is when all nodes have the same degree, links, remaining links or cycles. As of the maximum entropy, each network has a limit which is a function of number of network nodes.

Another picture of entropy is derived directly from the original definition of entropy by Boltzmann. The work in [12] proposes to compute the network entropy based on the topology configurations of the network. By varying the network parameters, several network configurations are generated. These configurations are constrained by the network model. For instance, in a random network model, the total number of configurations is limited by the wiring probability. The results in [12] assert that their approach can be extended to other models and can estimate the network deviation from equilibrium.

The entropy based on the number of configuration is defined as

$$S(W, \mathbf{p}) = - \sum_{k=1}^W P(k) \ln(P(k)) \quad (2)$$

Where the probability in this case is the probability of finding a given configuration.

```

MaxDegreeEntropyGraph(n)
  for i → 1 to nodesNumber
    network.addNode(i)
  processedNodes = 0
  for i := 1 to ⌊n/2⌋
    for j := 1 to (n - 2*processedNodes - 1)
      Network.nodes(n-i).connect(network.nodes(n-i-j))
    processedNodes++
  return network

```

Figure 1 The MDE algorithm.

#### 4. Topology selection to maximize network entropy

Our study strives to maximize the entropy of networks by carefully selecting the models to generate such networks. We studied and suggested models and methods that would explore the best network topologies that can maximize network entropies. We explore both degree and cyclic entropies.

##### 4.1. The topology of maximum degree entropy

To find the best topology of a network with the maximum degree entropy (MDE), we propose and employ a deterministic simple algorithm that manipulates and adjusts the degrees of network nodes to produce the required topology of the network.

Given a network of  $N$  nodes, the maximum degree of any node in this network is  $N-1$ . And the possible degree for any node is between 1 and  $N-1$ . Since we have  $N$  nodes and  $N-1$  different degrees, then according to Pigeonhole principle, any distribution of the possible degrees to the nodes would end up with at least two nodes mapped to the same degree. We propose the algorithm shown in Figure 1 to find the topology of a network that has a uniform degree distribution (or almost uniform). Since we are using a uniform distribution, then at least one of the degrees must occur twice.

The algorithm described generates a network that contains all the possible degrees, with at least two nodes with the exact same degree. First, all the nodes are generated with no links attached to them. Then, at each stage  $i$  of the algorithm we connect any node to  $N-i$  nodes excluding the nodes that we already used in the previous stages. After repeating this process for  $\lfloor N/2 \rfloor$  times, we would guarantee to generate the whole graph (because of symmetry) and that the nodes would be mapped to one of the available degrees from 1 to  $N-1$ , except for two nodes that will have the degree  $\lfloor N/2 \rfloor$ . This generated topology will have the highest degree entropy since almost all the nodes have the same degree probability. Samples of generated networks are shown in Figure 2.

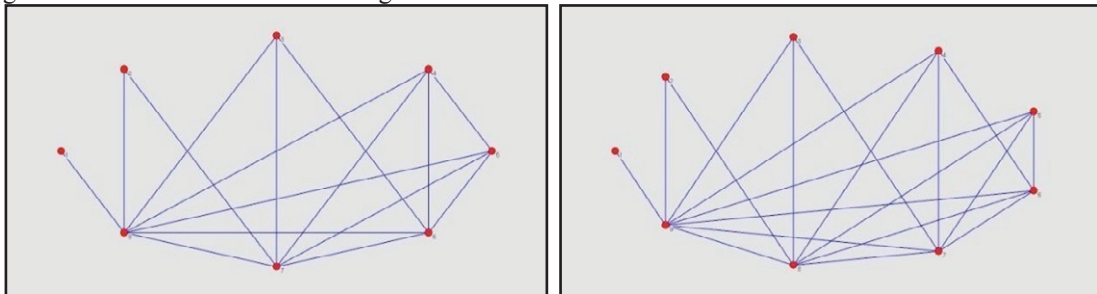


Figure 2. The maximum possible degree entropy A. size 8, B. Size 9

The degree entropy of networks generated using the algorithm in Figure 1, and can be calculated mathematically in terms of only the number of nodes ( $N$ ) in the network, using eq (1-2). The summation of the degrees of the nodes will be normalized by dividing them by the  $N$  for the nodes with degrees that occurred only once and dividing them by  $N/2$  for the nodes with degrees that occurred twice. The degree entropy is then calculated by:

$$\text{degree entropy} = -\frac{(N-2)}{N} \ln \left[ \frac{1}{N} \right] - \frac{2}{N} \ln \left[ \frac{2}{N} \right] \quad (3)$$

#### 4.2 The topology of maximum cyclic entropy

One of the hard problems (NP-Complete) is to generate a network that satisfies a certain cycle's distribution. The number of cycles in a graph is usually exponential in terms of the number of nodes in the graph ( $N$ ). Moreover, what adds to the complexity of such an algorithm is that not all the cycles' distributions that are randomly generated represent actual graphs. This is hardly due to the fact that the number of cycles of size  $k$  depends on the number of cycles of size  $l$  up to  $k-1$  that exists in the graph. Adding one link/node to the graph, can increment all the cycle size but with a non-uniformly distributed values. Furthermore, unlike degree distribution, removing one link can cause a large change to the number of cycles of different sizes, while it only changes the degrees of two nodes (reducing them by one.) Hence, in this work we propose using a genetic algorithm as described in Figure 3 to explore/generate the topologies that exhibit maximum cyclic entropy.

The algorithm starts by generating an initial generation of random networks [1]. A random network is a well known type of complex networks that can be generated using the algorithm described in Figure 4. After generating the initial generation/population of random networks, the algorithm enters a loop that starts with calculating the fitness values of the genes (where each gene represents a random network). The fitness value in our algorithm is calculated by computing the cycle's distribution for each random network using Johnsons' algorithm [13]. The cycle's distribution is randomized by dividing them by the sum of all cycles. The resulted normalized distribution is used to calculate the entropy using the equation:

$$S(W, \mathbf{p}) = -\sum_{k=1}^W P(k) \ln(P(k))$$

Where the probability in this case is the probability of finding a given network configuration /connections.

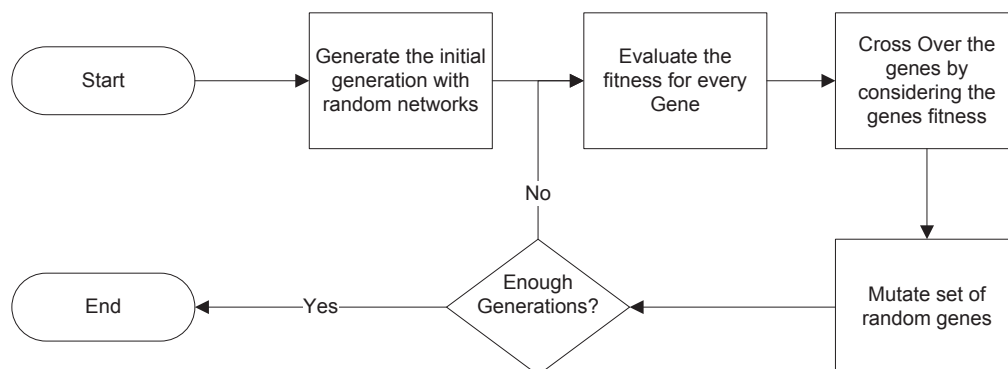


Figure 3 The genetic MCE algorithm.

Using the fitness of the generated random networks, the algorithm applies a cross over operation on the old generated networks by exchanging random links between two different networks to generate a new random network, hopefully with a better fitness value. The larger the value of the fitness of the network, the higher the probability that it will be used in the cross over operation. Then the algorithm picks a random number of the previously generated random networks to mutate them by flipping the state of some links (i.e., leaving or removing links in the network.) The loop stops when it reaches the number of generations desired and set by the user. By applying such an algorithm, we look forward to generate random networks with the maximum cyclic entropy (MCE.)

The genetic algorithm discussed was used to find the topology of a network that gives the maximum cyclic entropy for size 7. The result is shown in Figure 5A.

As mentioned before, a systems' entropy is related to the network stability. Maximizing the network stability must lead to maximizing the network entropy. Some previous literatures suggested that the star network is the most stable network (See Figure 5B). However, this type of networks does not contain any cycles. This will lead to cyclic entropy of zero. The genetic algorithm we proposed creates a network that is so similar to the star network. The difference is the ring that passes through all the non-central nodes. We call this topology the cyclic star network. This network exhibits almost a uniform distribution. Another feature of this network is it can be expanded to fit any number of nodes.

```

randomNetwork(nodesNumber, probability)
  for i → 1 to nodesNumber
    network.addNode(i)
  for i → 1 to nodesNumber
    node1 = network.getNode(i)
    for j → 1 to nodesNumber
      if i ≠ j
        node2 = network.getNode(j)
        r = randomFraction()
        if (r ≤ probability)
          network.connect(node1, node2)
  return network

```

Figure 4 The random network generation algorithm.

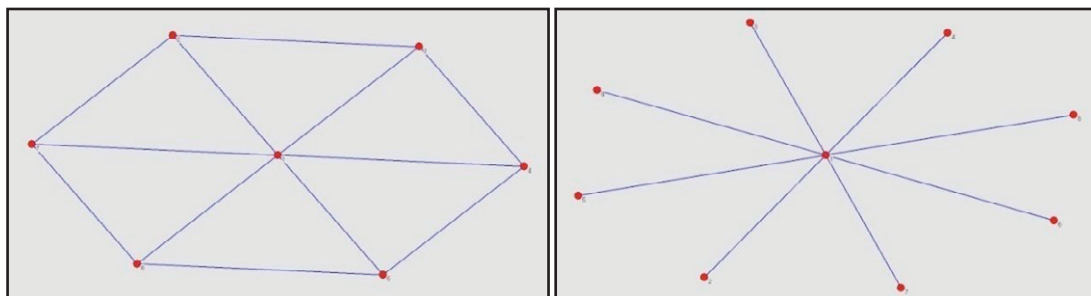


Figure 5 A. The cyclic star network of size 7. B. The star network of 9 nodes

The uniform cycle's distribution can be justified as follows. The network will be partitioned to a set of  $N-1$  triangles. Those triangles share with their direct neighbors only one link. Every triangle represents a cycle of length 3. Now, extracting two adjacent triangles will create a cycle of size 4. The question here is, how many two adjacent triangles do we have in such a network? Again it is  $N-1$ . That is because every two adjacent triangles have a distinct triangle. And we have  $N-1$  distinct triangles, thus we have  $N-1$  two adjacent triangles. This justification is applicable for any number of adjacent triangles. A set of  $d$  adjacent triangles represent a cycle with length  $(2+d)$ . There is one exception. Cycles of length  $N-1$  will have a one extra occurrence. This occurrence is represented by the ring that passes all the non-central nodes.

An analytical mathematical expression to find the cyclic entropy of this type of networks is shown here. First we need to calculate the summation of all cycles of the different sizes. We have  $N-3$  cycles that are of size  $N-1$  and we have only one cycle that is of size  $N$ .

Then the sum can be calculated using the following:

$$Sum = (N-1)(N-3) + N \quad (4)$$

Now, we can normalize the cycle sizes. The result will be as follows:

$$P(L = N-1) = \frac{N}{(N-1)(N-3) + N} \quad (5)$$

$$P(L \neq N-1) = \frac{N-1}{(N-1)(N-3) + N} \quad (6)$$

Then, we can calculate the cyclic entropy using the following equation:

$$\text{Cyclic Entropy} = -\frac{N}{(N-1)(N-3) + N} \ln \left[ \frac{N}{(N-1)(N-3) + N} \right] - \frac{(N-3)(N-1)}{(N-1)(N-3) + N} \ln \left[ \frac{N-1}{(N-1)(N-3) + N} \right] \quad (7)$$

## 5. Discussion and Conclusions

We have compared two models to generate networks with maximum entropies, one based on degrees and another based on cycles. The entropies of those generated networks are compared in this section. We developed a genetic algorithm to help in identifying such networks, where degree entropy and cyclic entropy were used as our fitness values.

To confirm the findings in section 4, a genetic algorithm (GA) is run for the generated configurations by the deterministic MDE algorithm (Figure 1). Apparently, the MDE algorithm results are better than GA as shown in Figure 6. For small size networks, the results are the similar.

The purpose of the genetic algorithm is to find the topology of a network with the maximum degree entropy. However, intuitively, there is only one possible cycle's distribution with higher cyclic entropy, which is the uniform cycle's distribution. For instance, in size 8, the genetic algorithm generated an irregular network (Figure 7) that has entropy larger than suggested cyclic star network. However, the difference is within statistical error, 0.001.



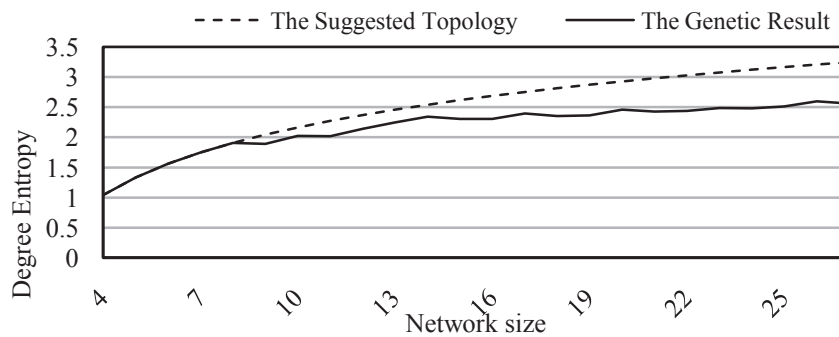


Figure 6. Comparison between GA and MDE algorithm for the evaluation degree entropy.

Based on GA, it is reasonable to conclude that a cyclic star network will give the maximum cyclic entropy. On the other hand, based on MDE algorithm, certainly a uniform distribution should result in the maximum degree entropy as expected.

## 6. Future Work

We intend to expand our models and algorithms to identify complex networks with maximum cyclic entropy. Creating a network with a specific cycle's distribution is a challenging problem; hence we plan to develop a new GA algorithm to help in generating such complex networks. We will also investigate other network models and configurations that are similar to the star network configuration and we believe that those models would also generate networks with optimal/maximum entropies.

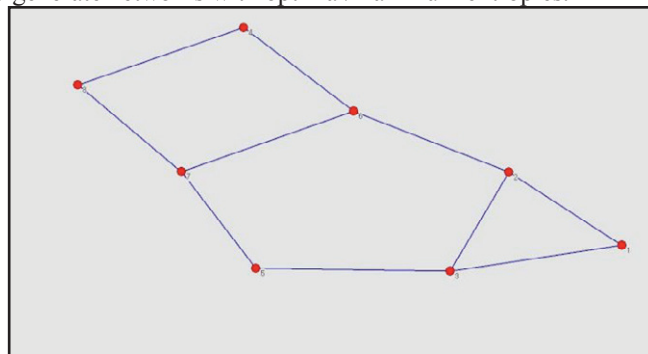


Figure 7 The irregular network that has uniform cycle's distribution

## References

- [1] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, "Characterization of Complex Networks: A Survey of measurements," presented at the Instituto de Física de São Carlos, Universidade de São Paulo, Brazil., 2006.
- [2] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, January 2002.
- [3] S. H. Strogatz, "Exploring Complex Networks," *NATURE*, vol. 410, pp. 268-276, 2001.



- [4] K. A. Mahdi, M. Safar, I. Sorkhoh, and A. Kassem, "Cycle-Based versus Degree-based Classification of Social Networks," *Journal of Digital Information Management*, vol. 7, 2009.
- [5] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378-382, 2000.
- [6] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1997.
- [7] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*: University of Illinois Press, 1975.
- [8] K. Mahdi, M. Safar, and I. Sorkhoh, "Entropy of Robust Social Networks," in *IADIS International Conference e-Society*, Algarve, Portugal, 2008.
- [9] M. Safar, K. Mahdi, and I. Sorkhoh, "Maximum Entropy of Fully Connected Social Network," in *IADIS International Conference Web Based Communities*, Amsterdam, Holland, 2008.
- [10] D. P. Feldman, "A brief introduction to: Information Theory, Excess Entropy and Computational Mechanics," Department of Physics, University of California, July, 1997.
- [11] R. V. Sole and S. Valverde *Information Theory of Complex Networks: On Evolution and Architectural Constrains*. Lect. Notes Phys. 2004 650 p.189-207.
- [12] Li Ji, B.-H. Wang, W.-X. Wang and T. Zhou *Network Entropy bases on Topology COnfiguration and Its Computation to Random Network* *Chin. Phys. Lett.* 2008. 25(11)
- [13] D. Johnson, *Finding All the Elementary Circuits of a Directed Graph*. *SIAM Journal on Computing*, 1975. 4(1): p. 77-84.
- [14] P. Erdős, and A. Rényi, *On random graphs*. *Publicationes Mathematicae*, 1959. 6: p. 290-297.
- [15] R.E. Tarjan, *Enumeration of the Elementary Circuits of a Directed Graph*. 1972, Cornell University.
- [16] H. Liu and J. Wang, *A new way to enumerate cycles in graph*, in *Proceedings of the Advanced Int'l Conference on Telecommunications and Int'l Conference on Internet and Web Applications and Services*. 2006, IEEE Computer Society.
- [17] P. Erdos, and A. Rényi, *On the evolution of random graphs*. *Publ. Math. Inst. Hungar.Acad. Sci*, 1960. 5: p. 17-61.
- [18] P. Erdos, and A. Rényi, *On the strenght of connectedness of a random graph*. *Acta Mathematica Scientia Hungary.*, 1961. 12: p. 261-267.
- [19] E. Marinari and G. Semerjian, *On the number of circuits in random graphs*. *Journal of Statistical Mechanics: Theory and Experiment*, 2006.
- [20] J.S. Yedidia, W.T. Freeman, and Y. Weiss, *Understanding belief propagation and its generalizations*, in *Exploring artificial intelligence in the new millennium*. 2003, Morgan Kaufmann Publishers Inc. p. 239-269.
- [21] P.D.L. Rios, S. Lise, and A. Pelizzola, *Bethe approximation for self-interacting lattice trees*. *EPL (Europhysics Letters)*, 2001. 53(2): p. 176-182.